

# Inverse Quantum Chemistry: Concepts and Strategies for Rational Compound Design

Thomas Weymuth and Markus Reiher<sup>1</sup>

ETH Zürich, Laboratorium für Physikalische Chemie,  
Vladimir-Prelog-Weg 2, 8093 Zürich, Switzerland

## Abstract

The rational design of molecules and materials is becoming more and more important. With the advent of powerful computer systems and sophisticated algorithms, quantum chemistry plays a decisive role in the design process. While traditional quantum chemical approaches predict the properties of a predefined molecular structure, the goal of inverse quantum chemistry is to find a structure featuring one or more desired properties. Herein, we review inverse quantum chemical approaches proposed so far and discuss their advantages as well as their weaknesses.

Date: March 9, 2014

Status: printed in *Int. J. Quantum Chem.* 114 **2014** 823–837

---

<sup>1</sup>corresponding author; e-mail: markus.reiher@phys.chem.ethz.ch

# 1 Introduction

The last decades have witnessed the fast-paced development of a wide range of quantum chemical methods [1]. Examples are highly accurate but also computationally demanding wave-function-based approaches like coupled cluster and configuration interaction [2] and density functional theory [3]. The latter allows for the description of large molecular systems consisting of hundreds of atoms — usually at reduced accuracy, however. Today, these computational approaches are firmly established in chemistry, and they are crucial for basically all areas of chemical research.

Underlying all these methods is the time-independent Schrödinger equation,

$$\hat{H}\Psi = E\Psi, \quad (1)$$

where  $\Psi$  is the sought-for wave function characterizing the system under investigation,  $E$  the energy associated with this wave function, and  $\hat{H}$  is the Hamiltonian operator. In nonrelativistic quantum chemistry, the Hamiltonian of an assembly of (point-like) atomic nuclei and electrons is given by

$$\hat{H} = -\frac{1}{2} \sum_I \frac{1}{M_I} \Delta_I - \frac{1}{2} \sum_i \Delta_i - \sum_{i,I} \frac{Z_I}{r_{iI}} + \sum_{i,j>i} \frac{1}{r_{ij}} + \sum_{I,J>I} \frac{Z_I Z_J}{r_{IJ}} \quad (2)$$

(in Hartree atomic units, in which the elementary charge, the mass of the electron, the reduced Planck constant, and  $4\pi\epsilon_0$  take a value of one — used throughout this work, unless otherwise stated). The indices  $I, J$  and  $i, j$  run over all nuclei and electrons, respectively.  $M_I$  denotes the mass of nucleus  $I$ ,  $Z_I$  its charge number, and  $r_{ij}$  is the spatial distance between particles  $i$  and  $j$ . The Laplacian is defined as

$$\Delta_i = \frac{\partial^2}{\partial x_i^2} + \frac{\partial^2}{\partial y_i^2} + \frac{\partial^2}{\partial z_i^2}. \quad (3)$$

The first two terms in Eq. (2) are the kinetic energy operators of the nuclei and electrons, while the last three terms describe the electron–nucleus attraction, the electron–electron repulsion, and the nucleus–nucleus repulsion, respectively. The rather complicated Hamiltonian of Eq. (2) is usually simplified by invoking the well-known Born–Oppenheimer

approximation, in which the nuclear coordinates are treated as (fixed) parameters. We may then define the electronic Hamiltonian

$$\hat{H}_{\text{el}} = -\frac{1}{2} \sum_i \Delta_i - \sum_{i,I} \frac{Z_I}{r_{iI}} + \sum_{i,j>i} \frac{1}{r_{ij}} + \sum_{I,J>I} \frac{Z_I Z_J}{r_{IJ}}, \quad (4)$$

and the corresponding electronic Schrödinger (eigenvalue) equation

$$\hat{H}_{\text{el}} \Psi_{\text{el}} = E_{\text{el}} \Psi_{\text{el}}. \quad (5)$$

The molecular structure, defined as an assembly of atomic nuclei fixed in space, is a direct consequence of this approximation.

For a given assembly of atomic nuclei and electrons, the nonrelativistic electronic Hamiltonian is unequivocally defined. The methods mentioned in the beginning aim at an approximate solution of the electronic Schrödinger equation. In principle, all observables and molecular properties of interest can then be calculated from the wave function, which contains all information that can be known about the system [4]. Therefore, quantum chemical methods provide us with a wealth of information about a certain system, but only *after* this very system has been specified in terms of the nuclear framework. For the design of new functional molecules or materials one encounters the reverse situation, where a desired property is known and a molecule featuring this very property is searched for. Given the “forward” direction from structure to property, extensive screening of structures is necessary to find a molecule with a predefined property (one example for this approach is drug design). It is thus highly desirable to develop computational approaches which would render such time- and cost-intensive trial-and-error procedures obsolete. In order to achieve this goal, the direction needs to be reversed, i.e., from property to structure. Such approaches are called “inverse approaches”.

Inverse problems do not only occur in the field of chemistry, but play an important role in diverse areas such as geophysics (where the composition of the Earth’s mantle is inferred from gravimetric measurements [5]), medical imaging (where, for example, attempts are made to clarify the internal structure of tissues by ultrasonic pressure waves [6]) and computer vision (where three-dimensional data from lower-dimensional images is constructed [7]). Accordingly, the theory of inverse problems constitutes a whole branch of

mathematics with important contributions made in the first half of the 20th century by Russian researchers (see Ref. [8] for an extensive bibliography). Clearly, since inverse problems represent an important aspect in many fields of research, the study of inverse problems in quantum chemistry may well benefit from developments made in other fields [9].

As we will see, the greatest challenge an inverse chemical application faces is the huge size of chemical space, i.e., the space of molecules accessible by contemporary synthetic protocols. It has been estimated that this size is between  $10^{20}$  and  $10^{24}$  molecules [10]. However, an estimation of the total size of chemical space highly depends on the assumptions made; for example, the number of molecules having up to 30 carbon, nitrogen, oxygen, and sulfur atoms has been estimated to be larger than  $10^{60}$  [11], while the number of proteins that could theoretically exist is roughly  $10^{390}$  [12] (for an average size of 300 amino acid residues per protein). This might be the reason for the fact that inverse quantum chemical approaches have only emerged during the past 20 years, although the basic idea has been presented much earlier [13]. It is the purpose of this work to review all existing approaches in this field and to analyze their potential for future applications.

We will first review in section 2 the mathematical concepts developed for inverse Sturm–Liouville problems, where important results emerged as early as 1929 and strongly influenced research carried out between 1950 and 1970 on inverse quantum scattering. Then, we review inverse techniques which have either already found widespread application in rational compound design or exhibit a great potential. We discuss the inverse perturbation analysis in section 3 and the use of model equations in section 4. Next, we move on to more recent approaches, which rely on sophisticated sampling and optimization techniques, such as the optimization of wave functions in section 5, quantitative structure–activity relationships in section 6, and the optimization of solid state compounds in sections 7.1 and 7.2. We review two very recent approaches, namely the linear combination of atom-centered potentials in section 7.3 and alchemical potentials in section 7.4. Finally, we also highlight developments from our research group in sections 8 and 9 followed by general conclusions in section 10.

## 2 Inverse Spectral Theory

Inverse spectral theory is the branch of mathematics which studies what can be deduced about the structure of (differential) operators given that some of their properties are known [14]. In quantum chemistry, this translates to the question whether it is possible to reconstruct the Hamiltonian from a predefined set of eigenvalues.

The Soviet–Armenian physicist Viktor Ambarzumian [15] investigated to which extent a differential equation is defined by its eigenvalue spectrum [16]. Basic inverse questions—although not explicitly stated as such—have been studied earlier in a purely mathematical context (see, for example, Ref. [17]). In fact, Lord Rayleigh briefly discussed an inverse problem already in 1877 in his work about acoustic waves [18]. Moreover, Ambarzumian was well aware of the fact that the ability of reconstructing the Hamiltonian from its eigenvalue spectrum would yield great insight into the structure and properties of matter [16]. He studied the simple case of an oscillating string, mathematically modeled by a differential equation of second order,

$$-\frac{d^2\phi(x)}{dx^2} + q(x)\phi(x) = \alpha\phi(x); \quad 0 \leq x \leq 1 \quad (6)$$

with the boundary conditions

$$\frac{d\phi(0)}{dx} = \frac{d\phi(1)}{dx} = 0, \quad (7)$$

where  $\phi(x)$  describes the oscillation as a function of position  $x$ .  $q(x)$  is a continuous function, and  $\alpha$  is a constant. He showed that the series of eigenvalues

$$\alpha_n = n^2\pi^2; \quad n \in \mathbb{N} \quad (8)$$

is only consistent with the differential equation, Eq. (6), if  $q(x) \equiv 0$ . It is easily seen that Eq. (6) resembles the Schrödinger equation for a particle in a one-dimensional box if we take  $\alpha$  as proportional to the energy eigenvalue and  $q(x)$  as a position-dependent potential. Note, however, that the boundary conditions employed by Ambarzumian are different from the ones of the particle in a one-dimensional box. Ambarzumian studied so-called natural boundary conditions, where the first derivative of the function must vanish

at the box boundaries, whereas for the ‘particle in the box’ model the function itself must be zero at the boundaries.

From the results of Ambarzumian, it follows that we can reconstruct the potential  $q(x)$  — and concomitantly the full Hamiltonian — if we know the (complete) set of energy eigenvalues (see also Ref. [14]). For example, we can choose a potential  $q$  such that the resulting Hamiltonian features an eigenvalue spectrum of  $\mu^2\pi^2, n^2\pi^2$  with  $n \in \mathbb{N}$ ,  $1 < \mu < 2$  and  $n > \mu$  (i.e.,  $n = 1$  is not possible and the lowest eigenvalue is shifted from  $n = 1$  to  $n' = \mu$ ). Thus, we can selectively shift only the lowest eigenvalue while leaving all others unchanged. In this case, the (unnormalized) wave function corresponding to the lowest eigenvalue  $\mu^2\pi^2$  can be shown [14] to be given by

$$\phi_\mu(x) = \frac{\sin(\pi x)}{\left[ \frac{\sin(\mu(1-x)) - \sin(\mu x)}{\sin(\mu)}, \sin(\pi x) \right]}, \quad (9)$$

where the denominator is defined as

$$[f(x), g(x)] = f(x) \frac{dg(x)}{dx} - \frac{df(x)}{dx} g(x), \quad (10)$$

the so-called Wronskian of  $f$  and  $g$ . The function given in Eq. (9) is depicted in Fig. 1 a) for  $\mu^2 = 3$ , and compared to the (regular) ground state wave function of an electron in a one-dimensional box, i.e., for  $q'(x) = 0$ .

The corresponding potential can easily be constructed numerically by inserting Eq. (9) into the Schrödinger equation for a particle in a one-dimensional box and solving for  $q$ . It is plotted in Fig. 1 b).

Several different methods to construct a Hamiltonian from a given energy spectrum have been developed. However, in 1946, the Swedish mathematician Göran Borg generalized Ambarzumian’s work [19]. In his seminal paper, Borg showed that in the general case, one *cannot* determine the function  $q(x)$  from only one eigenvalue spectrum. Instead, one also needs the spectrum of a “complementary” eigenvalue problem (e.g., the same differential equation with different boundary conditions) in order to fully determine the corresponding differential operator. Thus, in the general case we cannot deduce the form

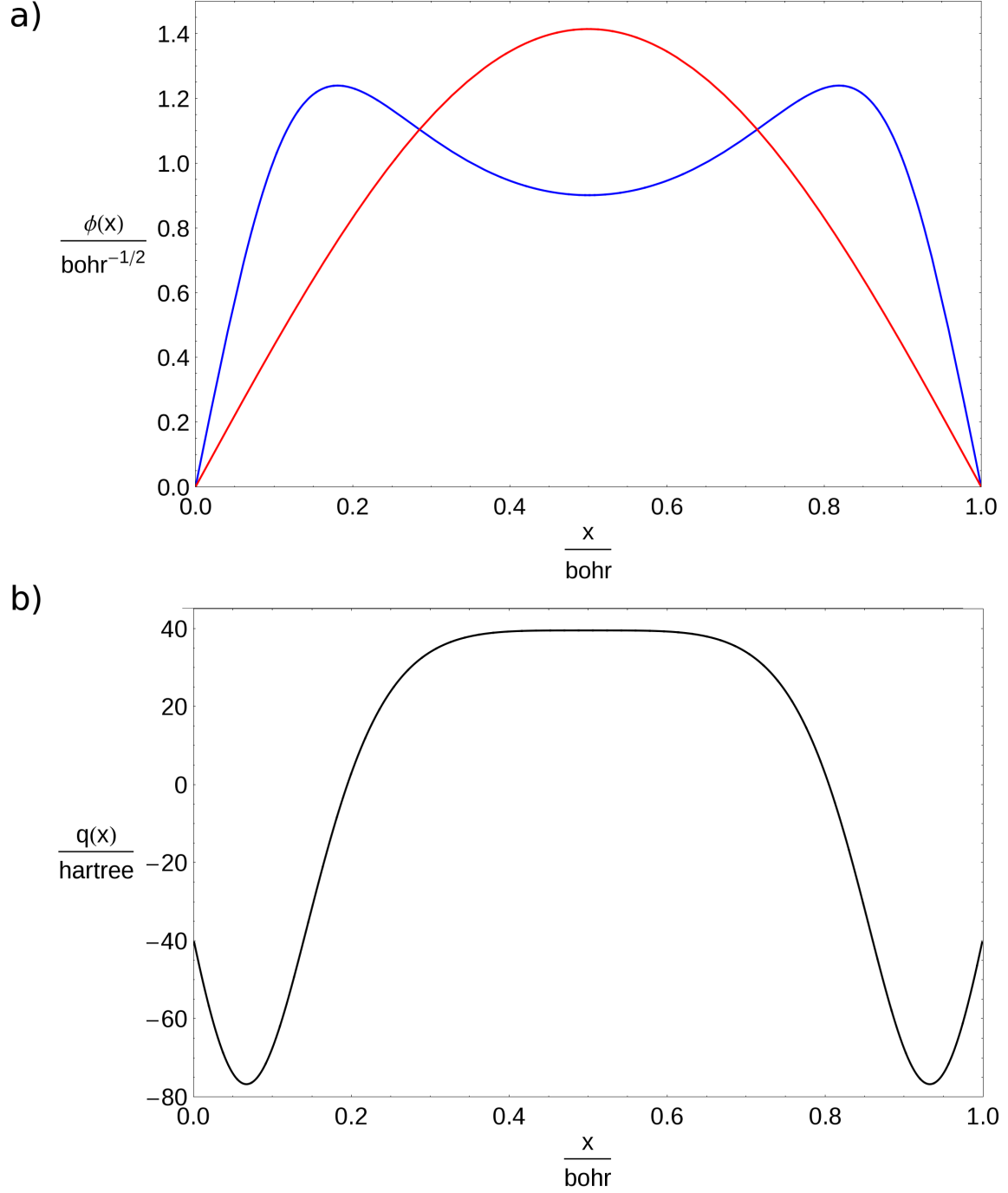


Figure 1: a) Normalized wave functions for an electron in a one-dimensional box of length 1 bohr. Red: ground state wave function when no potential energy term is present in the Hamiltonian. Blue: ground state wave function if an additional potential is introduced such that the lowest eigenvalue is shifted from  $\pi^2$  to  $3\pi^2$  while all others remain unchanged. b) Potential  $q(x)$  producing the wave function of Eq. (9) with  $\mu^2 = 3$ .

of the Hamiltonian by predefining a suitable set of energy eigenvalues, as is done in the above example. It was pointed out that such an inversion is in general not possible, since different Hamiltonians (representing different molecules) can actually feature comparable properties [20, 21]. Even if it would be possible to strictly invert the Schrödinger equation in the most general way, it can well be imagined that among the infinitely many different molecules, there will always be molecules that feature almost identical values for a given property. One can thus intuitively understand that for a given application, several molecules can exhibit the correct property. Therefore, different approaches have to be devised to tackle this difficult problem. In the following, we will review the most important of these approaches.

### 3 Inverse Perturbation Analysis

In molecular spectroscopy, one is often interested in determining the potential leading to a certain energy spectrum as accurately as possible. Nowadays, there exist many different approaches to accomplish this task [22–25], and there are also specialized computer programs available (see, e.g., Ref. [26]). For the sake of brevity, we shall not extensively dwell here on the details of all these approaches. However, we would like to highlight one method, which is very general in its scope of use and can also serve as motivation for the development of more advanced inverse approaches, as we shall see.

In 1975, Kosman and Hinze developed a method they called Inverse Perturbation Analysis (IPA) which allows one to calculate accurate potential energy curves [27]. They started from the one-dimensional Schrödinger equation for a diatomic molecule (note, however, that the method itself is not limited to diatomics),

$$\left(-\frac{1}{2\mu}\frac{d^2}{dr^2} + U(r)\right) R(r) = ER(r), \quad (11)$$

where  $\mu$  is the reduced mass of the molecule under consideration,  $r$  the internuclear distance,  $R(r)$  the nuclear wave function, and  $E$  the corresponding energy eigenvalue.  $U(r)$  is the potential energy function, which is usually not known exactly. However, many approx-



imate potentials are known, which are often very close to the unknown exact potential like the well-known Rydberg–Klein–Rees potential. One of the central ideas of IPA is now to write the exact potential as a sum of a known approximate potential  $U^{(0)}$  and a (small) correction term  $\Delta U$ ,

$$U(r) = U^{(0)}(r) + \Delta U(r). \quad (12)$$

If the approximate potential is a good approximation to the exact curve, the correction term is expected to be small, and can consequently be searched for with an iterative perturbative approach, where  $\Delta U$  is the perturbation. First, Eq. (11) is solved numerically for the wave functions  $R^{(0)}$  and energy eigenvalues  $E^{(0)}$  with the approximate potential  $U^{(0)}$ . Since the true energy eigenvalues  $E$  can be obtained experimentally by spectroscopic techniques, one can calculate an energy correction

$$\Delta E = E - E^{(0)}, \quad (13)$$

which can be approximated by the first-order energy correction

$$\Delta E = \langle R^{(0)}(r) | \Delta U(r) | R^{(0)}(r) \rangle. \quad (14)$$

If one expands the potential correction into a set of predefined suitable basis functions,  $f_i(r)$ ,

$$\Delta U(r) = \sum_i c_i f_i(r), \quad (15)$$

then the energy correction becomes

$$\Delta E = \sum_i c_i \langle R^{(0)}(r) | f_i(r) | R^{(0)}(r) \rangle, \quad (16)$$

which can be solved by standard methods to obtain the best set of coefficients  $c_i$ . Once this set is obtained, one can use the coefficients to improve on the representation of the actual potential energy curve by employing Eq. (12). With this improved potential, one can iterate the procedure outlined above to obtain an even better potential function. Thus, one can stepwise improve the description of the potential energy until the energy correction term  $\Delta E$  is below a given threshold.

In their original paper, Kosman and Hinze successfully applied their method to HgH. However, they found that their results heavily depend on the particular choice of basis functions  $f_i(r)$ . While Kosman and Hinze employed global polynomials (i.e., polynomials defined over the entire definition range of  $r$ ,  $-\infty \leq r \leq \infty$ ), Hamilton and coworkers relied on local Gaussian functions and found accurate and rapidly convergent results several years later [28]. Still, Hamilton and coworkers experienced some numerical instabilities (the results were very sensitive to changes in the parameters of the basis functions). These instabilities are a characteristic of inverse problems: they are typically ill-conditioned. In our case, the behavior of the potential energy function  $U(r)$  in regions where the wave function  $R(r)$  is zero is not determined, as can be seen by inspection of Eq. (14), and, therefore, the solution of this equation is not unique. We may recall that it can be mathematically proven that a single energy eigenvalue spectrum is in general not sufficient to determine the underlying potential uniquely. To remedy this problem of numerical instability, Wu and Zhang proposed to solve Eq. (14) by means of a singular value decomposition and found that their approach is fast, accurate and numerically stable [29].

We thus have a working method to produce accurate molecular potential energy functions, given the exact energy spectrum is known and a good approximate potential can be found as a starting point. This is certainly very interesting for the inverse point of view. However, we are still confronted with two major problems: first, we are usually not that much interested in a molecule featuring a given energy eigenvalue spectrum, but rather in a molecule with a specific property (such as a specific dipole moment). These properties can, however, usually not be directly related to a specific energy spectrum. Furthermore, even if we could somehow overcome this problem and come up with a certain molecular potential, it would not automatically be clear how the underlying molecular structure would be composed of atomic nuclei and electrons. However, this problem could be solved in a quite elegant fashion by employing some sort of atom-based basis functions, such that each basis function would represent one (abstract) atom—the actual type of the atom would then be determined by the final expansion coefficient  $c_i$ . In the above optimization procedure, one would, of course, not only have to search for the expansion coefficients, but also for the position at which the basis functions are centered (which enter as a

parameter in these local functions and denote the position of the nuclei in space), unless one uses a comparatively large number of such basis functions, such that they can simply be distributed uniformly in space. At a position where no nuclei should be, we can expect that the respective expansion coefficient is close to zero. In fact, such approaches have been emerging during the past few years, see sections 7.3 and 7.4.

Even though there appears to be no general solution of the first problem mentioned above, we shall see in the sections to come that we might find other ways of directly relating the molecular potential to properties other than the energy eigenvalue spectrum, thereby making this approach a very promising one. We close this section with a reference to work by Politzer and Murray [30,31] who elaborated on the central role of the electrostatic potential created by atomic nuclei and electrons as the determinant of molecular properties and reactive behavior.

## 4 Model Equations

A completely different and very interesting approach is the development of (simple) model equations. This technique has been successfully employed by Marder *et al.* in 1991. They describe possibilities to increase the first electronic hyperpolarizability of conjugated organic molecules [32]. We illustrate the underlying design approach in the following, but note that the design of molecules with optimal optoelectronic properties has been studied by many groups [33–35].

If a molecule is placed in an external electric field, its charge density is polarized. When the external electric field has only a comparatively small effect on the electronic structure of a given molecule, the polarization  $\mathbf{P}$  can be written as a power series

$$\mathbf{P} = \alpha \mathbf{E} + \beta \mathbf{E}^2 + \gamma \mathbf{E}^3 + \mathcal{O}(\mathbf{E}^4), \quad (17)$$

where  $\mathbf{E}$  is the external electric field, and the constants  $\alpha$ ,  $\beta$ , and  $\gamma$  are the polarizability, first hyperpolarizability, and second hyperpolarizability, respectively. We note in passing that in the general case, the (hyper-)polarizabilities are not scalar quantities but instead

described by tensors. However, for the following discussion, our approximation does not constitute any loss of generality. Since the electric field constitutes only a minor perturbation of the molecule, perturbation theory is a suitable method to calculate  $\alpha$ ,  $\beta$ , and higher order corrections (in cases where the wave length of irradiation is much larger than the size of the molecule, the perturbation Hamiltonian describes the interaction of the molecular dipole moment with the external electric field). The first hyperpolarizability  $\beta$  can thus be obtained by correcting the wave function to second order in the electric field [36]. The resulting expression is rather complicated, involving sums over all occupied orbitals. However, it is often observed that the first hyperpolarizability is dominated by contributions from only the first few excited states. This motivates a simplification by taking only the ground state “g” and the first excited state “e” into account. The resulting two-state approximation reads [36],

$$\beta \propto \frac{X_{ge}^2 (X_{gg} - X_{ee})}{(E_g - E_e)^2}, \quad (18)$$

with the matrix elements defined as

$$X_{ij} = \langle \Psi_i | \mathbf{r} | \Psi_j \rangle; \quad i, j \in \{e, g\}. \quad (19)$$

The influence of the three quantities  $X_{ge}^2$ ,  $X_{gg} - X_{ee}$ , and  $1/(E_g - E_e)^2$  on the first hyperpolarizability can be studied at a simple two-orbital system — following Ref. [36], for which we may write the wave function as a linear combination of these two orbitals,

$$\Psi = c_1 \phi_1 + c_2 \phi_2. \quad (20)$$

This linear combination is to be understood as a quantum-mechanical model wave function constructed from orbitals  $\phi_i$  defined on different parts of the complete molecule (this model is in the spirit of classical molecular orbital theory like Hückel theory, i.e., a proper construction of a wave function as an antisymmetrized Hartree product and the consideration of spin are left aside).

We now define the energy expectation values

$$\langle \phi_1 | H | \phi_1 \rangle = \Delta, \quad (21)$$

and

$$\langle \phi_2 | H | \phi_2 \rangle = -\Delta, \quad (22)$$

such that the energy difference between the two orbitals  $\phi_1$  and  $\phi_2$  is  $2\Delta$ . The coupling strength between the two orbitals is then defined as

$$\langle \phi_1 | H | \phi_2 \rangle = t. \quad (23)$$

With these definitions, we can write the time-independent Schrödinger equation in matrix form as

$$\begin{pmatrix} \Delta - E & t \\ t & -\Delta - E \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \quad (24)$$

For this system of homogeneous equations to have a nontrivial solution, i.e.,  $c_1, c_2 \neq 0$ , the determinant of the  $2 \times 2$  matrix must be zero, which yields an equation to determine the energy  $E$ . With the two solutions

$$E_{\pm} = \pm \sqrt{\Delta^2 + t^2}. \quad (25)$$

and the first equality of Eq. (24),

$$(\Delta - E_{\pm}) c_1^{(\pm)} + t c_2^{(\pm)} = 0, \quad (26)$$

we can establish the relation

$$\frac{c_1^{(\pm)}}{c_2^{(\pm)}} = \frac{E_{\pm} - \Delta}{t} = \pm \sqrt{1 + \left(\frac{\Delta}{t}\right)^2} - \frac{\Delta}{t}. \quad (27)$$

For each of the two energy eigenvalues, we get a separate set of coefficients  $c_i$ , such that we obtain two wave functions,

$$\Psi_g = \Psi_- = c_1^{(-)} \phi_1 + c_2^{(-)} \phi_2, \quad (28)$$

and

$$\Psi_e = \Psi_+ = c_1^{(+)} \phi_1 + c_2^{(+)} \phi_2. \quad (29)$$

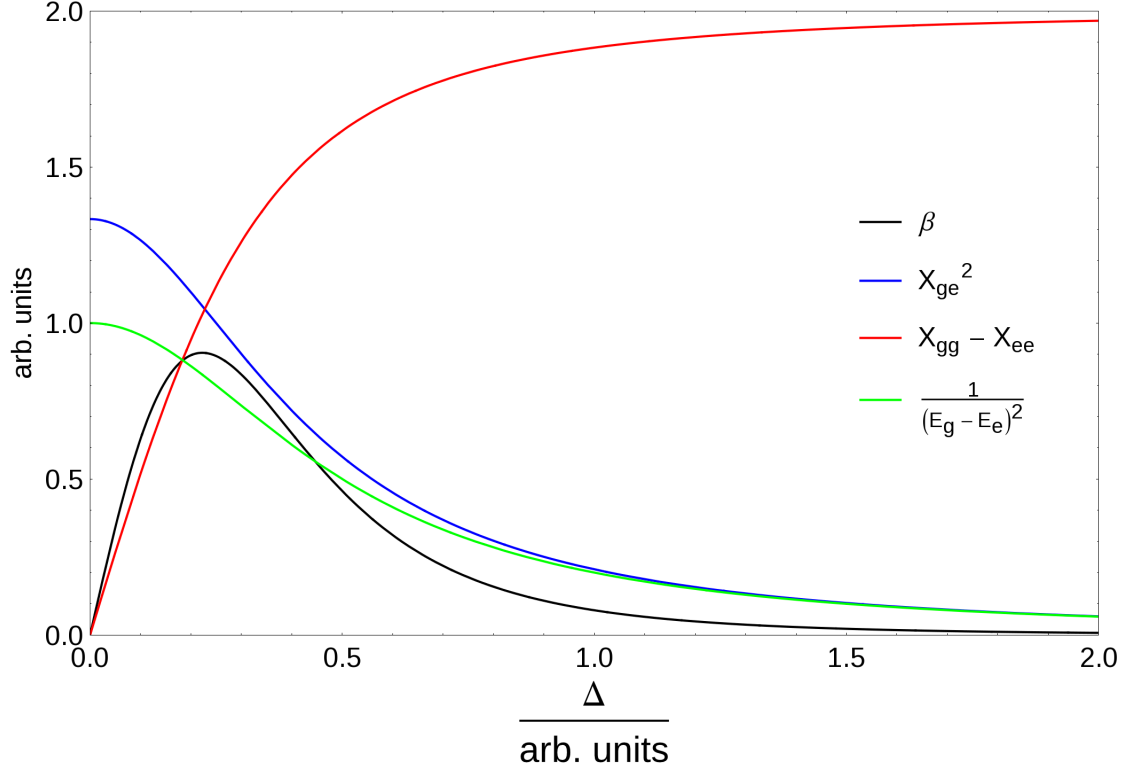


Figure 2: The first hyperpolarizability  $\beta$  of a simple two-orbital system — calculated from its components  $X_{ge}^2$ ,  $X_{gg} - X_{ee}$ , and  $1/(E_g - E_e)^2$  according to Eq. (18) — as a function of  $\Delta$ , where  $2\Delta$  is the energy difference between the two orbital energies. In this figure, the overlap between the two orbitals as well as the coupling strength  $t$  were set to 0.5.

Expressions for the individual coefficients can be obtained from the normalization condition, i.e.,

$$\langle \Psi_{\pm} | \Psi_{\pm} \rangle = 1. \quad (30)$$

We can thus calculate expressions for  $X_{ge}^2$ ,  $X_{gg} - X_{ee}$ , and  $1/(E_g - E_e)^2$  as well as for  $\beta$  as a function of  $\Delta$ . These functions are given in Fig. 2. We see that the combination of the individual expressions leads to a maximum of  $\beta$  at about  $\Delta = 0.25$  when the overlap between  $\phi_1$  and  $\phi_2$  and the coupling strength  $t$  are set to 0.5. The task is now to find a molecule corresponding to these values, which is by far not trivial.

In their publication, Marder *et al.* also exploited the mathematical expression for the first hyperpolarizability  $\beta$  given in Eq. (18), but for a somewhat more sophisticated model

system consisting of four orbitals (corresponding to donor, acceptor, and bridge orbitals) [32]. Marder *et al.* could relate the above expressions to actual molecular building blocks, in such a way that they were able to come up with specific design strategies to construct the desired molecules [32].

Unfortunately, the deduction of an actual molecular (sub-)structure is not straightforward at all, but requires often a considerable amount of chemical knowledge and intuition. Especially this last fact is a notable drawback, limiting the applicability of this method to a rather narrow subspace of chemical compound space. Therefore, this approach suffers from the same fundamental problem as the QSPR techniques (see below).

## 5 Optimized Wave Functions

This severe limitation has long been recognized [37]. In 1996, Kuhn and Beratan proposed a new strategy in order to overcome this problem, which can be regarded as a first truly inverse approach in rational compound design [37]. They suggested to search for an optimal molecular structure by optimizing the corresponding mathematical objects describing it, i.e., the Hamiltonian and its wave function. The authors also point out that this inverse problem is ill-conditioned as not enough information is present in order to carry out the optimization (recall Section 3). This problem can be lifted by introducing additional constraints. For example, bonding in molecules is local by nature, leading to characteristic patterns in the Hamiltonian matrix. This can be conveniently illustrated by the example of the simple Hückel molecular orbital method [38, 39], where a linear conjugated hydrocarbon chain corresponds to a tridiagonal Hamiltonian matrix due to a nearest-neighbor approximation. The overlap of basis functions located at distant atoms is in fact usually negligibly small, and thus also in more advanced electronic structure theories the Hamiltonian matrix is usually sparse. With the Hückel method, for a linear chain of three orbitals centered at  $x_i = \{0, 1, 2\}$ , we can write the Hamiltonian in matrix

form as

$$H = \begin{pmatrix} \alpha_1 & \beta_1 & 0 \\ \beta_1 & \alpha_2 & \beta_2 \\ 0 & \beta_2 & \alpha_3 \end{pmatrix}, \quad (31)$$

where the  $\alpha_i$  represent the orbital energies and the  $\beta_i$  the coupling between adjacent orbitals. Since we are free in our choice of energy origin, we can arbitrarily set one of the  $\alpha_i$  to zero. Furthermore, we can also choose any suitable unit for the energy, which allows us to set one of the  $\beta_i$  to  $-1$ . With this, we can write

$$H \rightarrow H' = \begin{pmatrix} 0 & -1 & 0 \\ -1 & \alpha'_2 & \beta'_2 \\ 0 & \beta'_2 & \alpha'_3 \end{pmatrix}. \quad (32)$$

The task is now to find values for the remaining parameters such that a given property is optimized. As an example, Kuhn and Beratan chose the electric dipole operator, the matrix elements of which can be written as

$$\mu_{km} = - \sum_{i=1}^3 x_i c_i^{(k)} c_i^{(m)}, \quad (33)$$

if overlap contributions are neglected ( $c_i^{(k)}$  is the expansion coefficient of the  $i$ -th orbital in wave function  $k$  as in section 4) and optimized the transition dipole moment  $\mu_{23} = c_2^{(2)} c_3^{(3)} + 2c_3^{(2)} c_3^{(3)}$ . Within our simple approach,  $\mu_{23}$  is a function of only three variables, namely  $\alpha'_2$ ,  $\alpha'_3$ , and  $\beta'_2$ . If we arbitrarily set  $\beta'_2 = \beta_1 = -1$ , we can plot  $\mu_{23}$  as a function of only  $\alpha'_2$  and  $\alpha'_3$ , as shown in Fig. 3. It is straightforward to optimize  $\mu_{23}$  with respect to these two parameters. For  $\alpha'_2 \geq -5$ , one finds that  $\mu_{23}$  is minimized for  $\alpha'_2 = -5.00$  and  $\alpha'_3 = 0.00$ , which yields  $\mu_{23} = -0.97$  [37]. The global minimum, however, appears to be at  $\alpha'_2 = -4848.59$  and  $\alpha'_3 = 0.00$ . With these values, one finds  $\mu_{23} = -1.00$ .

In their pioneering study, Kuhn and Beratan proposed also to optimize not only the elements in the Hamiltonian matrix, but also the coefficients of the wave function (while the energy eigenvalues were held fixed). They applied their method to a few very simple “toy” examples like the one given above in order to demonstrate the general concept. However, for many years after its publication, the method could not be applied to find an



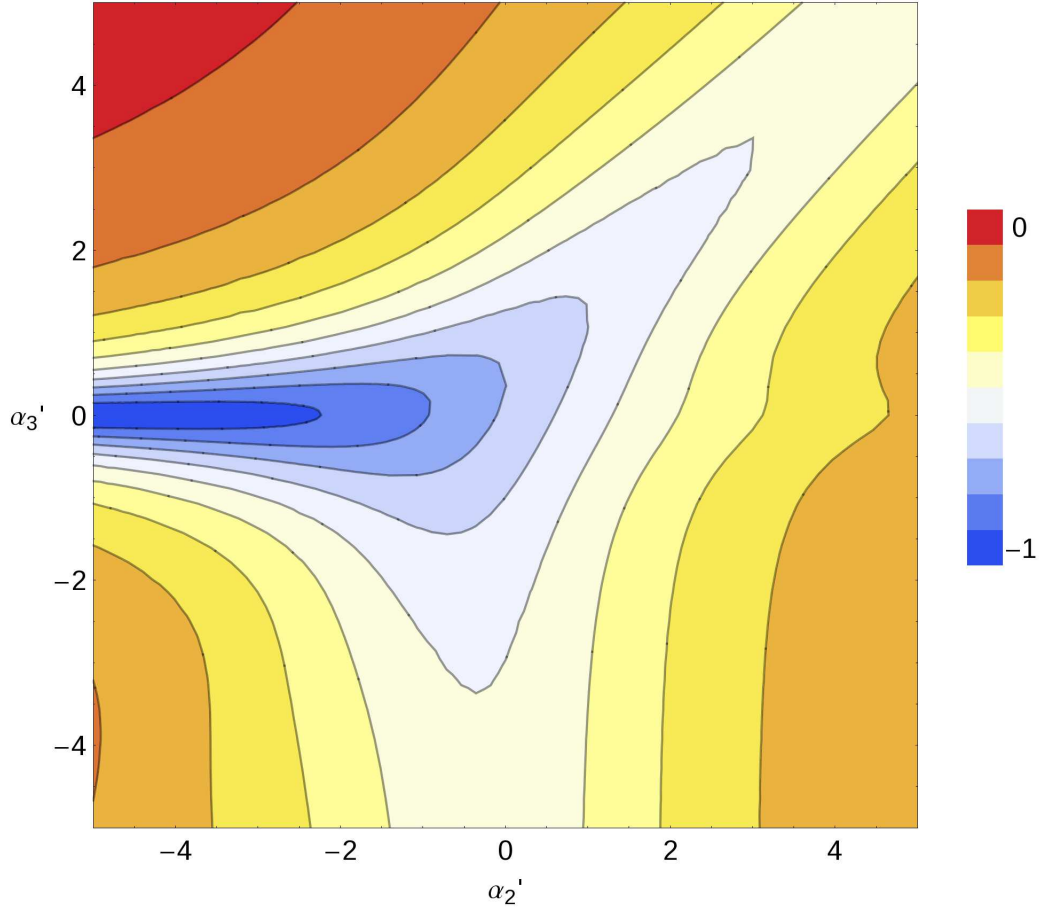


Figure 3: Dependence of the transition dipole moment  $O_{23}$  on the two parameters  $\alpha'_2$  and  $\alpha'_3$ , respectively. All data are given in arbitrary units.

actual molecule exhibiting a desired property, as was hoped by the authors. The problem was not so much the sheer size of chemical compound space, but rather the fact that it is extremely difficult to construct a molecule from its Hamiltonian matrix or wave function [40]. However, in 2006, Beratan and coworkers established a general framework which can circumvent this problem (see Section 7.3).

## 6 Quantitative Structure–Activity Relationships

Quantitative Structure–Activity Relationships (QSAR) and the closely related Quantitative Structure–Property Relationships (QSPR) are central for design attempts in chemistry, biology, and materials research [41–47]. The origins of QSAR trace back to as early as 1863, when Cros described a relation between the toxicity of primary aliphatic alcohols and their solubility in water [48]. However, the actual mechanistic basis of contemporary QSAR approaches was laid in 1964 by Hansch and Fujita with the development of the linear Hansch equation [49], which was based on the famous work of Hammett, who proposed an equation to relate equilibrium constants of a particular reaction involving benzene derivatives to some structural parameter, which depends on the particular type of benzene substituents, and a reaction parameter, which depends on the type of reaction under study [50, 51].

The basic assumption of QSPR is that every physical, chemical, and biological property depends in a systematic way on the underlying molecular structure. The functional form of this dependence is then searched for by trying to establish a model relating a given property to one or more descriptive parameters of the structure (so called descriptors). In a typical QSPR study, a training set of molecular structures and corresponding properties is taken as input. The descriptors are then usually calculated directly from the molecular structure. As descriptors, diverse quantities such as the molecular weight, the number of ring systems, the dipole moment, or the solvent-accessible molecular surface can be employed [44]. Even though one could also use descriptors which are obtained experimentally, it is advantageous to employ descriptors which can be obtained directly from

the (three-dimensional) molecular structure without the need of time-consuming measurements. Once all desired descriptors have been obtained, a statistical analysis of these data is carried out in order to find a mathematical relation between the descriptors and the target property. In some cases, a simple linear function can already be used to relate one descriptor to a target property, while in other cases more complex, nonlinear, multivariate expressions might be necessary. In contemporary QSPR studies, sophisticated methods such as artificial neural networks and support vector machines are employed in order to establish a functional dependence between properties and descriptors [46]. Once such a dependence has been established, one can easily predict the desired properties for arbitrary molecules (although one has to be careful when predicting a property for a molecule which is very different from any molecule of the original training set [47]).

Let us illustrate this by a very simple example. Given a set of nine trigonal-bipyramidal molybdenum complexes shown in the top right corner of Fig. 4, we would like to find an expression giving the Mo–N<sub>2</sub> binding energy as a function of some descriptor(s). As the Mo–N bond length can be expected to depend on the binding energy, it appears natural to choose this bond length as a descriptor.

Fig. 4 shows the binding energies as a function of the Mo–N bond length [52]. In a first approximation, this dependence can be described by a linear function. With a least-squares fit, the binding energy  $E$  can be obtained from the Mo–N bond length  $d$  as

$$E(d) = 8.31 \text{ kJ mol}^{-1} \text{ pm}^{-1} d + 1818.73 \text{ kJ mol}^{-1}. \quad (34)$$

As one can see from Fig. 4, this relation is overly simplified. Even though the Mo–N bond length can be used as a first, rough criterion to estimate the bond strength, it is certainly not the only factor affecting this quantity. This can be seen from two complexes in Fig. 4, which have almost the same bond length (namely, about 198.5 pm), but largely differ in their binding energy. Our calculated expression above would predict the same binding energy for both complexes, thus overestimating the value for one complex and underestimating it for the other. A descriptor which could be better suited to judge the Mo–N<sub>2</sub> binding energy would be the stretching frequency of the Mo–N or N–N bonds. With such an approach, a relationship between the N–N stretching frequency and the

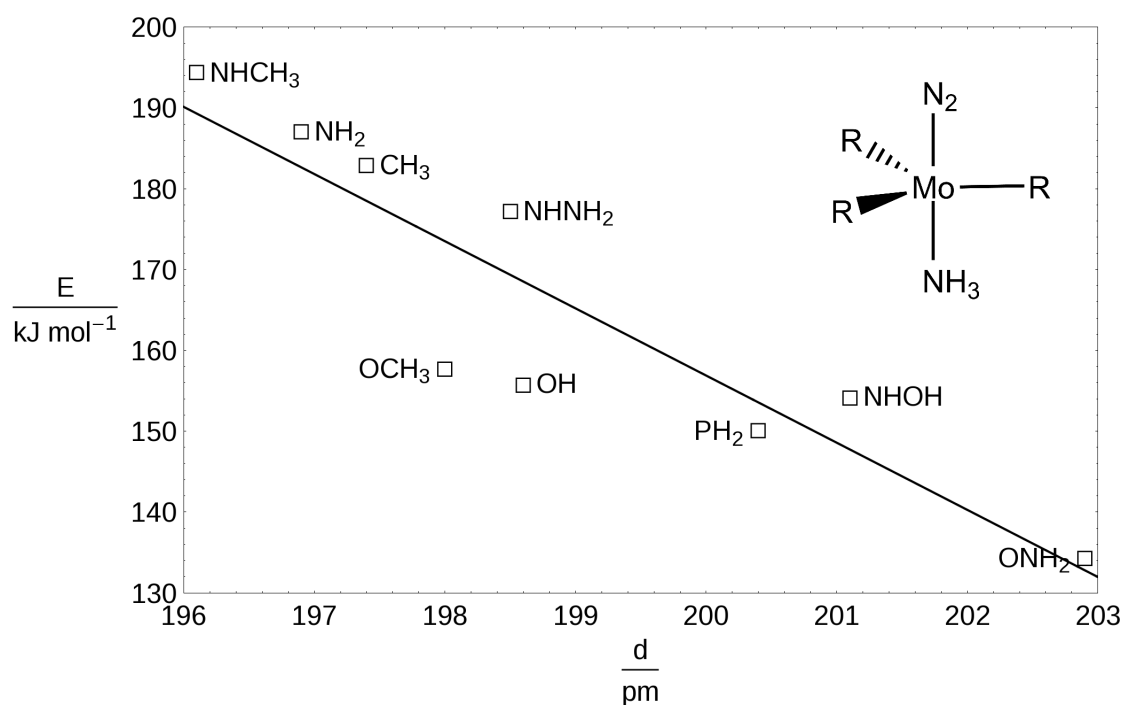


Figure 4: Dependence of the Mo-N<sub>2</sub> binding energy  $E$  on the Mo-N bond length  $d$  of the trigonal-bipyramidal molybdenum complexes with the general structure shown in the top right corner. Calculated data for the molecules are taken from Ref. [114] and are shown as black squares (labeled with the individual residues R). A linear least-squares regression is given by the black line.

degree of activation of dinitrogen for a series of Schrock-type nitrogen fixing catalysts was established [53].

Of course, the development of a robust QSPR for a given quantity is not trivial, and cannot be simply automatized. In particular, it is often not *a priori* clear which descriptors should be chosen. Furthermore, a given QSPR is usually only valid for a comparatively small subset of chemical compound space, namely for molecules which are very similar to each other in terms of their molecular structures. Therefore, these techniques are only of limited applicability for a truly inverse approach, to which the entire chemical compound space should be accessible. Moreover, even though it is easy to identify the descriptor values

necessary for a given desired property once a mathematical relation between this property and its descriptors has been found, the construction of a molecule corresponding to a given set of descriptors is not straightforward at all. An approach to tackle this problem is denoted as inverse QSAR [54,55]. In this approach, one aims at constructing a library of chemical compounds similar to a given lead compound (which is known to exhibit some favorable property). In this context, chemical similarity is defined as the Euclidean distance of two molecules in the space spanned by all descriptors used. Therefore, the more similar the descriptor values are for two compounds, the more chemically similar they are expected to be. The generation of new molecules is accomplished by connecting different predefined building blocks with each other in order to obtain a chemically reasonable structure. Then, the descriptors of this new structure are computed. They can be used in an optimization algorithm such as simulated annealing in order to find structures which exhibit a predefined similarity to the lead compound. One can thus identify further compounds worth to be closely investigated without the need of screening huge molecular libraries.

## 7 Advanced Sampling Techniques

Powerful computer systems and clever optimization algorithms (e.g., simulated annealing [56], genetic algorithms [57], and Monte Carlo sampling [58]) allow us to sample a comparatively huge part of chemical space. The application of such algorithms is the most important approach in rational compound design. Therefore, we shall also review these approaches, although they are not strictly inverse, since they do not directly aim at predicting a structure featuring a predefined set of properties, but rather follow the more traditional approach where the properties of a predefined molecular structure are calculated.

## 7.1 Rational Design in Solid-State Chemistry

During the early 1990s, Schön and Jansen developed a sophisticated technique for synthesis planning [59,60]. They proposed a modular computer program, which searches chemical space by means of a global optimization technique and identifies promising structure candidates, that can subsequently be refined and analyzed. After some external boundary conditions (e.g., initial number of atoms, external pressure, etc.) have been specified, the algorithm sets up an elementary cell. The unit cell vectors (length and orientation) are chosen randomly, as are the positions and composition of atoms. Then, all these parameters are globally optimized employing a fast but not very accurate electronic-structure method. The final result is then subjected to a more accurate refinement. The global optimization is carried out multiple times, starting with different random simulation cells. The resulting collection of promising compounds can then be analyzed in terms of their composition and physical properties. Jansen and coworkers have applied their methodology to a range of example systems as, for example, noble gas mixtures [61] and binary ionic compounds [62].

## 7.2 Inverse Band Structure Approach

A similar approach has been mentioned in 1994 by Werner *et al.* who proposed to tailor a band structure reflecting the desired properties and to subsequently search for a solid exhibiting this very band structure [63]. They termed this problem of finding a solid with a suitable band structure the “inverse band structure problem”. Unfortunately, they could not come up with a working solution to this problem. At this point, we should illustrate how complex this optimization problem is. Consider a comparatively simple system, such as the pseudo-binary alloy  $A_{0.25}B_{0.75}C$  with a unit cell of 128 lattice sites (64 cation sites, occupied by either A or B, and 64 anion sites, all of which are occupied by C). Then, the total number of possible configurations is given by the binomial coefficient of 64 and 16, i.e., it is on the order of  $10^{14}$  [64]. Needless to say that exhaustive enumeration of such a big search space is simply not possible. However, only five years later, Franceschetti

and Zunger published a computational procedure which successfully addresses the inverse band structure problem [64]. Similarly to the approach by Schön and Jansen, they start with an elementary cell of a given configuration. They define an atomic configuration in terms of a vector

$$\sigma = (S_1, S_2, \dots, S_N), \quad (35)$$

where  $S_i$  is the atom type at lattice site  $i$ . A given property  $P$  (such as the band gap) is a function of the atomic configuration. The goal of the optimization is to find a configuration with one or several predefined target properties  $P^{(\text{target})}$ , i.e., one attempts to minimize the function

$$O(\sigma) = \sum_{\alpha} \omega_{\alpha} |P_{\alpha}(\sigma) - P_{\alpha}^{(\text{target})}|, \quad (36)$$

where  $\omega_{\alpha}$  is the weight assigned to the property  $P_{\alpha}$ . In the original approach by Franceschetti and Zunger, this function is minimized by a simulated annealing algorithm. Atomic configurations are generated by Monte Carlo moves (e.g., changing the type of a given atom). Even though this algorithm is able to find the global minimum significantly faster than it would be possible with exhaustive enumeration, it still requires sampling of several thousand atomic configurations. It is therefore very important to have a fast method at hand to evaluate the properties of a given atomic configuration. Franceschetti and Zunger relied on a valence-force-field method to optimize the positions of the atoms of a given atomic configuration [65], and utilized a semiempirical Hamiltonian, where each atom is replaced by a pseudopotential [66], in order to evaluate the properties. Furthermore, they applied a special technique for the diagonalization of the Hamiltonian, which focuses on the few eigenvalues around the band gap [67].

Franceschetti and Zunger predicted [64] the configuration of an  $\text{Al}_{0.25}\text{Ga}_{0.75}\text{As}$  alloy with a maximum band gap. Recently, Zunger and coworkers studied a quaternary  $(\text{In}, \text{Ga})(\text{As}, \text{Sb})$  semiconductor [68]. The original approach has continuously been further developed and improved. For example, it is also possible to employ a genetic algorithm instead of simulated annealing [69, 70], which generally leads to a faster convergence.

We should note here that the *general* method of finding the optimal species with regard

to a given property in a subset of chemical space is also employed and further developed by many other groups, although they do not specifically aim at an inverse approach. For example, Nørskov and coworkers have identified the 20 most stable four-component alloys of 32 transition, noble and simple metals in both face-centered cubic and body-centered cubic structures [71]. For this, they searched a subspace of no less than  $192 \times 10^{16}$  compounds by means of an evolutionary algorithm [71]. Given the limited space here, we refer the reader to Refs. [72–78] for further examples of computational materials design and to Refs. [79,80] for an example of environment-enhanced reactivity provided by inverse solvent design of a condensed phase reaction.

The approach by Franceschetti and Zunger is specifically tailored for the design of solid-state materials, and its extension to the design of isolated molecules is not necessarily straightforward. In the above procedure, the total number of atoms is defined from the outset, as a certain unit cell with a given number of lattice sites is defined. However, for an isolated molecule, this approach is not valid. One would therefore have to abandon this constraint and introduce different ones. For example, by means of simple valence rules, one could define the total number of atoms coordinating to a given atom (i.e., the number of nearest neighbors). Furthermore, in the design of solid-state materials one can apply periodic boundary conditions, and therefore in principle study a crystal of infinite size. For an isolated molecule, however, this is not possible, and one therefore encounters the problem of unsaturated valencies.

### 7.3 Linear Combination of Atom-Centered Potentials

One of the more daunting problems in the field of rational design is the fact that individual molecules represent discrete points in chemical space, whereas most optimization methods are developed to work with continuous functions. In 2006, Beratan, Yang and coworkers addressed this problem [81]. They started from the well-known fact that within the framework of density functional theory any molecular system is characterized by the external potential  $v(\mathbf{r})$  [82], which represents the attraction between electrons and nu-



clei, and the total number of electrons. They then treated this external potential as a continuous variable to be optimized. Yang *et al.* even presented a functional of external potentials together with a corresponding variational principle [83]. One major issue with this approach is that not all external potentials correspond to a molecule, even though all molecules map unequivocally to a given  $v(\mathbf{r})$ . An external potential which corresponds to a chemical structure is called C-representable. Beratan, Yang and coworkers very elegantly solved this problem of C-representability by expanding the potential  $v(\mathbf{r})$  into a set of atomic potentials [81],

$$v(\mathbf{r}) = \sum_{A, \mathbf{R}} b_A^{(\mathbf{R})} v_A^{(\mathbf{R})}(\mathbf{r}), \quad (37)$$

where  $b_A^{(\mathbf{R})}$  is the expansion coefficient of the potential term  $v_A^{(\mathbf{R})}(\mathbf{r})$ , which denotes the potential of atom  $A$  at position  $\mathbf{R}$ . This approach has been called “linear combination of atomic potentials” (LCAP), and it constrains the external potential to be necessarily C-representable. The optimization problem then reduces to finding the best set of expansion coefficients for these atomic potentials. It is important to note that not only individual atoms can be represented by such atomic potentials but a collection of such potentials can be contracted to represent e.g., a functional group,

$$v_A^{(\mathbf{R})}(\mathbf{r}) = \sum_B b_B^{(\mathbf{R})} v_B^{(\mathbf{R})}(\mathbf{r}), \quad (38)$$

and can be optimized as a whole. In order to correspond to an actual molecule, the coefficients  $b_A^{(\mathbf{R})}$  must be either 0 or 1 denoting either the absence or the presence of the atom or functional group, respectively, represented by  $v_A^{(\mathbf{R})}(\mathbf{r})$ . However, since the expansion coefficients are treated as continuous variables during the optimization, one can find non-integer values for them. In this case, rounding to the nearest integer is necessary. While this approach of continuous optimization is rather new in quantum chemistry, Zunger and coworker pointed out that Bendsøe and Kikuchi had already presented a similar approach in the context of materials design [84].

In their original publication [81], Beratan, Yang and coworkers investigated as a proof of concept a rather simple problem, constructed from two sites (fixed in space), for which they searched for the best potential such that the resulting hyperpolarizability is maximized. In

doing so, they started from a small library of six different chemical substituents modeled by different potentials, and investigated how the optimal potential is composed of these “atomic” potentials. They finally found that  $\text{H}_2\text{S}_2$  has the largest hyperpolarizability of all molecules present in the small chemical space considered (six substituents at two sites give a total of  $2^6 = 64$  possible molecules) which was in agreement with the results found by exhaustive enumeration. With the same methodology, Keinan *et al.* successfully optimized the first hyperpolarizability of porphyrin-based nonlinear optical materials [85]. d’Avezac and Zunger recently published results showing that the approach by Yang and coworkers is very efficient for optimizing the atomic configurations of alloys [86].

In recent years, the group of Yang continuously developed and improved the method further. In 2007, they implemented it into an AM1 semiempirical framework [87]. In 2008, a gradient-directed Monte Carlo approach, which is able to “jump” between chemical structures, and thus to overcome barriers between local minima, was presented [88]. The approach was also implemented within a tight-binding framework [89]. Furthermore, these authors studied different optimization procedures and found that a combination of their LCAP approach with a best-first search algorithm (BFS, an algorithm searching a graph by stepwise expanding the most promising node [90]) improves the performance of the optimization [91].

The LCAP and related approaches have reached a state of maturity which allows them to be applied to rather complex problems [85]. However, the position of the potentials is fixed in space (and only optimized in the course of a standard structure optimization), which restricts the space of chemical compounds studied. This bears the danger of overlooking a good candidate, simply because it is not in the sampled space. This problem could be remedied by employing an extremely large library and allowing potentials at many different sites. However, this would also dramatically increase the computational cost of the resulting optimization problem. Nevertheless, as computer systems become more and more powerful, the LCAP will certainly be applied to increasingly complex problems.

## 7.4 Alchemical Potentials

Another interesting approach to convert chemical space into a continuous space has been developed by von Lilienfeld *et al.* Starting again from the Hohenberg–Kohn theorem [82], von Lilienfeld *et al.* noted that any molecular property  $O$  can be written as a functional of the nuclear charge distribution  $Z(\mathbf{r})$  and as a function of the total electron number  $N_{\text{el}}$  [21], i.e.,  $O \equiv O[Z(\mathbf{r})](N_{\text{el}})$ . Pictorially speaking, molecules are superpositions of electron and nuclear charge distributions. The electron density, however, is (up to a constant factor) fully determined by the total electron number and the external potential, which itself is a functional of the nuclear charge distribution. Rational compound design can thus be described as an optimization problem within the space spanned by nuclei and electrons, which leads us to the minimization problem [21]

$$\min_{N_{\text{el}}, Z(\mathbf{r})} |O[Z(\mathbf{r})](N_{\text{el}}) - O_0|^2, \quad (39)$$

where  $O_0$  is the desired target property. A continuous optimization in the space of electrons and nuclei implies fractional electronic and nuclear charges. At this point, we should note that some controversy concerning fractional electrons emerged some years ago and is still not fully resolved (see, e.g., Ref. [92]). In practice, in von Lilienfeld’s approach fractional atomic and electronic charges are generated by means of a suitably parametrized effective core potential (ECP) [93]. Therefore, there are no explicit fractional electrons occurring in the computation, but they are rather implicitly dealt with in the ECP.

In order to efficiently search the space spanned by electrons and nuclei, von Lilienfeld *et al.* worked out a mathematical expression for the derivative of the total electronic energies with respect to the nuclear charge distribution [21]. This has subsequently become known as the nuclear chemical potential  $\mu_n(\mathbf{r})$  (in analogy to the electronic chemical potential), which, for a discrete nuclear charge distribution and perturbation theory to the first order, is found to be the electrostatic field  $E^{(1)}(\mathbf{r})$ ,

$$\mu_n(\mathbf{r}) \approx E^{(1)}(\mathbf{r}) = - \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' + \sum_I \frac{Z_I}{|\mathbf{r} - \mathbf{r}_I|}, \quad (40)$$

where the integration is over the entire space. The nuclear chemical potential is a function of space and is related to the proton affinity [21]. At the position of a nucleus, it measures

the tendency of this nucleus to “mutate” into a different type [21]. At these specific points, the nuclear chemical potential is usually called the alchemical potential [21].

Von Lilienfeld and Tuckerman demanded a rigorous description of chemical space [94]. They found that such a description is tightly connected to grand-canonical ensemble theory and, hence, could benefit from earlier developments. It was soon recognized [94] that molecular properties such as the potential energy represent a state function within the space of  $Z(\mathbf{r})$  and  $N_{\text{el}}$ , and hence, any two points in this space can be connected by an order parameter, which is usually denoted as  $\lambda$ . This parameter can be used to interpolate between any two molecules (similar to the thermodynamic integration) and to study the variation of a given property as one molecule is continuously changed into the other [94–97]. In such studies, von Lilienfeld found that the path connecting any two molecules is generally not linear [96]. This makes the prediction of properties of unknown molecules based on results of known compounds very difficult. A general method of linearizing these connection paths has yet to be found, and von Lilienfeld even offers a prize (the equivalent of an ounce of gold) to the person finding a solution to this complicated problem [93].

So far, this approach has not been applied to many problems yet. As a proof of principle von Lilienfeld *et al.* [21] designed a nonpeptidic anticancer drug candidate by optimizing the atoms constituting the peptide unit in a peptidic starting structure (even though this peptidic structure already showed strong antitumoral activity, it would be cleaved *in vivo* by proteases, such that it can essentially not be employed as drug). In 2007, Marcon *et al.* showed that the energy of the highest occupied molecular orbital in benzene and its BN-doped derivatives can be precisely tuned [98].

## 8 Mode- and Intensity-Tracking

In this section we consider subspace iteration techniques that we have devised to target molecular vibrations, for which one can either provide a simple structural distortion as

a first guess (Mode-Tracking) or which pick up intensity in some spectroscopic set-up (Intensity-Tracking). Both Mode- and Intensity-Tracking are rather different from all of the above-mentioned approaches in the sense that they do not aim for the rational design of a molecular compound or material. However, they belong to the realm of inverse methods as they turn the usual procedure in quantum chemical calculations upside-down [99]. More specifically, Mode-Tracking [100] allows for the iterative, but direct optimization of qualitatively predefined (local or collective) molecular vibrations without the need of first calculating the entire Hessian matrix (as would be the case in the standard approach). Similarly, Intensity-Tracking enables one to iteratively converge a spectrum based on the most intense vibrations starting from intensity carrying structural distortions [101–104]. Thus, these two methods directly target selected molecular vibrations of some predefined characteristic and allow one to obtain these at lower computational cost than with the traditional methodologies.

The standard approach to calculate the  $k$ -th vibrational frequency of a molecule is to solve the harmonic eigenvalue problem

$$\mathbf{H}^{(\text{m})}\mathbf{q}_k = \lambda_k\mathbf{q}_k, \quad (41)$$

where the eigenvector  $\mathbf{q}_k$  is called the  $k$ -th normal mode and the eigenvalue  $\lambda_k$  is proportional to the square of the  $k$ -th vibrational angular frequency.  $\mathbf{H}^{(\text{m})}$  is the mass-weighted Hessian matrix (in Cartesian coordinates), the entries of which are given by [105]

$$H_{I\alpha,J\beta}^{(\text{m})} = \frac{1}{\sqrt{M_I M_J}} \left( \frac{\partial^2 E_{\text{el}}}{\partial r_{I\alpha} \partial r_{J\beta}} \right), \quad (42)$$

where  $M_I$  and  $\mathbf{r}_I$  denote the mass and spatial position of nucleus  $I$  ( $\alpha \in \{x, y, z\}$ ), respectively. In the following, we will abandon these two-component subscripts in favor of only one subscript ranging from 1 to  $3N$  written in lower case. The idea of Mode-Tracking is to selectively calculate certain normal modes without first computing the entire Hessian matrix, which is the most time-consuming step of a standard vibrational analysis [100]. It is usually straightforward to create a molecular distortion as an approximation for the desired vibration. One can therefore start with a guess mode

$$\mathbf{b} = \sum_{i=1}^{3N} b_i \mathbf{e}_i^{(\text{m})}, \quad (43)$$

where the sum runs over all Cartesian coordinates of the  $N$  nuclei,  $b_i$  is the displacement magnitude of the mode in the  $i$ -th component, and  $\mathbf{e}_i^{(\text{m})}$  are the mass-weighted nuclear Cartesian basis vectors. Taking  $\mathbf{b}$  as a first approximation to the desired normal mode  $\mathbf{q}_k$ , one can calculate the  $k$ -th element of the left-hand side of Eq. (41) as

$$\sigma_k = \left( \mathbf{H}^{(\text{m})} \mathbf{q} \right)_k = \sum_j \frac{1}{\sqrt{M_j M_k}} \frac{\partial^2 E_{\text{el}}}{\partial r_j \partial r_k} b_j = \frac{1}{\sqrt{M_k}} \frac{\partial^2 E_{\text{el}}}{\partial r_k \partial \mathbf{b}}. \quad (44)$$

The last equality means that  $\sigma_k$  can be calculated by computing the directional derivative of the nuclear gradient along the collective distortion  $\mathbf{b}$ , which can be conveniently done in a semi-numerical fashion [100]. In Mode-Tracking, the target mode is iteratively improved by a Davidson-type subspace iteration method [100]. In every iteration  $l$ , a new vector  $\mathbf{b}^{(l)}$  is generated, and concomitantly, also a new vector  $\boldsymbol{\sigma}^{(l)}$  is obtained. The  $i$  vectors generated up to iteration  $i$  can be assembled into the matrices  $\mathbf{B}^{(i)}$  and  $\boldsymbol{\Sigma}^{(i)}$ , respectively, to produce the matrix

$$\tilde{\mathbf{H}}^{(\text{m},i)} = \mathbf{B}^{(i)T} \mathbf{H}^{(\text{m})} \mathbf{B}^{(i)} = \mathbf{B}^{(i)T} \boldsymbol{\Sigma}^{(i)}, \quad (45)$$

i.e., it is not necessary to calculate the full Hessian matrix  $\mathbf{H}^{(\text{m})}$ . We then solve the small eigenvalue problem

$$\tilde{\mathbf{H}}^{(\text{m},i)} \mathbf{u}^{(i)} = \boldsymbol{\rho}^{(i)} \mathbf{u}^{(i)}, \quad (46)$$

where the matrices  $\boldsymbol{\rho}^{(i)}$  and  $\mathbf{u}^{(i)}$  contain the  $i$ -th approximations to the target eigenvalue  $\lambda_k$  and target normal mode  $\mathbf{q}_k$ , respectively. The best approximations, which we will denote as  $\rho_s^{(i)}$  and  $\mathbf{u}_s^{(i)}$ , are selected from  $\boldsymbol{\rho}^{(i)}$  and  $\mathbf{u}^{(i)}$ , and the residuum vector

$$\mathbf{r}_s^{(i)} = \sum_{l=1}^i u_{s,l}^{(i)} \left( \boldsymbol{\sigma}^{(l)} - \rho_s^{(i)} \mathbf{b}^{(l)} \right) \quad (47)$$

is calculated. From this residuum vector, a new basis vector  $\mathbf{b}^{(i+1)}$  is generated,

$$\mathbf{b}^{(i+1)} = \mathbf{X}^{(i)} \mathbf{r}_s^{(i)}, \quad (48)$$

where  $\mathbf{X}^{(i)}$  is a preconditioner. The convergence characteristics of the Mode-Tracking algorithm strongly depend not only on the initial guess mode, but also on the particular

choice of  $\mathbf{X}^{(i)}$ . However, experience shows that also very simple choices for the preconditioner (e.g., a unit matrix) facilitate fast convergence [106].

The Mode-Tracking approach has successfully been applied in studies of a large  $[(\text{Ph}_3\text{PAu})_6\text{C}]^{2+}$  cluster [107], molecular wires (i.e., long carbon chains) [108], adsorbate vibrations [109], and tetrameric methyl lactate clusters [110]. Furthermore, in 2008, Herrmann *et al.* integrated the Mode-Tracking methodology into a QM/MM framework [111]. Recently, Kovyrshin and Neugebauer [112, 113] presented a similar tracking methodology for the selective calculation of electronic excitations in the framework of time-dependent DFT.

Intensity-Tracking generally works according to the same subspace-iteration principle explained above for Mode-Tracking. However, the approach does not aim at refining a given user-defined vibration, but to yield those normal modes which are associated with highest intensities (for a given vibrational spectroscopy). Therefore, the most important difference to Mode-Tracking consists in the choice of the starting guess mode. A suitable guess mode is a molecular distortion that attains intensity. Such intensity carrying modes can generally be derived from an eigenvalue problem defined for the particular spectroscopy under consideration. This guess depends strongly on the kind of vibrational spectroscopy for which the highest-intensity modes are to be obtained. We have developed Intensity-Tracking for conventional infrared spectroscopy [103], Raman and Raman optical activity spectroscopy [104], as well as resonance Raman spectroscopy [102].

Mode- and Intensity-Tracking demonstrate that subspace iteration methods are a feasible way to iteratively solve an inverse problem—here, to optimize normal modes from pre-defined molecular distortions.

## 9 Gradient-driven Molecule Construction

The ultimate criterion for the stability of a given molecule is the Gibbs (or Helmholtz) free energy. However, in (chemical) processes involving energy changes much larger than the thermal energy  $RT$  the entropic (and temperature) contributions may be neglected as the

change in free energy is then dominated by the change in electronic energy. This situation simplifies the design problem as it is governed by the electronic structure of the reactants in the chemical process under study. The electronic energy difference is calculated from the electronic energies of the reactants. The latter are determined as stationary points on the Born–Oppenheimer potential energy surface  $E_{el}$  of Eq. (5), which is the mathematical object that relates all electronic energies calculated for fixed nuclear frameworks of a given number of atomic nuclei (and electrons). Consequently, a characteristic of stable reactants *and* of first-order transition structures is a vanishing first derivative of the electronic energy with respect to all nuclear coordinates, i.e., a vanishing geometry gradient. Therefore, the geometry gradient is the central property for determining the stability of a molecule. This fact is utilized in *Gradient-driven Molecule Construction* (GdMC) [114], a new concept that we have proposed for the rational design of functional molecules in 2012 [115].

Given some desired structural feature within a molecule or a molecular assembly, the length of the gradient on the atoms in the isolated structural feature, i.e., the length of the fragment’s gradient,

$$|\nabla_{\text{frag}} E_{\text{el}}| = \sqrt{\sum_{I \in \text{frag}} \left[ \left( \frac{\partial E_{\text{el}}}{\partial r_{I,x}} \right)^2 + \left( \frac{\partial E_{\text{el}}}{\partial r_{I,y}} \right)^2 + \left( \frac{\partial E_{\text{el}}}{\partial r_{I,z}} \right)^2 \right]}, \quad (49)$$

with  $r_{I,\alpha}$  ( $\alpha \in \{x, y, z\}$ ) being a nuclear Cartesian coordinate, is in general different from zero and not even small. The gradient components of this fragment can become vanishingly small upon embedding of the fragment in an appropriate molecular environment. Thus, the vanishing gradient criterion of GdMC can be exploited for the search of molecular scaffolds that fulfill a certain chemical purpose provided that this purpose can be pre-defined in terms of a molecular structure. For instance, one may ask the question which chelate ligand can stabilize a molecular fragment, such as a transition metal binding a small inert molecule, by designing the scaffold of the chelate ligand in such a way that the geometry gradient at each nucleus in the compound system vanishes.

The structure of the fragment may even be chosen in such a way that some bonds are elongated compared to the isolated reactants, from which the fragment is formed, so that the situation in the fragment represents some degree of activation with respect to the



isolated reactants. For example, when a dinitrogen molecule binds to a transition metal center, its triple bond might remain more or less intact upon coordination and thus the  $\text{N}\equiv\text{N}$  bond length hardly changes. On the other hand, upon binding of the dinitrogen ligand electron transfer can lead to an activation in such a way that a diazenoid double bond or even a hydrazinoid single bond is formed with the respective characteristic bond lengths [116, 116]. Hence, prototypical bond lengths and angles can be pre-defined as necessitated by the overall chemical process of which the species to be designed is an intermediate. Hence, a design may be carried out for each node of a reaction network and mutual stability dependencies may also need to be fulfilled resulting in a highly complex optimization problem. However, different optimization approaches are possible within GdMC, and we have explored some of them in our study on dinitrogen binding molybdenum complexes [114]. In a first approach, the Schrock complex, which is known to bind molecular nitrogen (see Fig. 5), was taken as an example of an existing nitrogen-binding complex that could be used as a known reference to validate the approach. In one optimization procedure, a range of model complexes was deduced from this compound in which a central  $\text{Mo-N}_2$  fragment was taken from the full Schrock complex but where the ligand sphere was modeled by the variable substituents  $\text{R}_1$  and  $\text{R}_2$ . Then, the atoms or functional groups for  $\text{R}_1$  and  $\text{R}_2$  were searched such that the geometry gradient on all atoms of the complex was minimal, leading to a stable complex featuring the predefined bond lengths mentioned above.

This procedure is illustrated in Fig. 5. In a first optimization cycle, different atoms (oxygen, nitrogen, carbon, and phosphorus) were placed at  $\text{R}_1$  and all valencies were saturated with hydrogen atoms. An analysis of the resulting geometry gradient revealed that oxygen is best suited for minimizing the geometry gradient. One can further reduce the gradient by varying  $\text{R}_2$  in further optimization cycles. We found that with an ethyl group for  $\text{R}_2$ , our model complex featured a gradient which was already very close to the reference gradient of the original Schrock complex.

So far we have taken only the first steps within the GdMC approach, but we expect that it can be turned into a black-box inverse as well as high-throughput screening tool for the

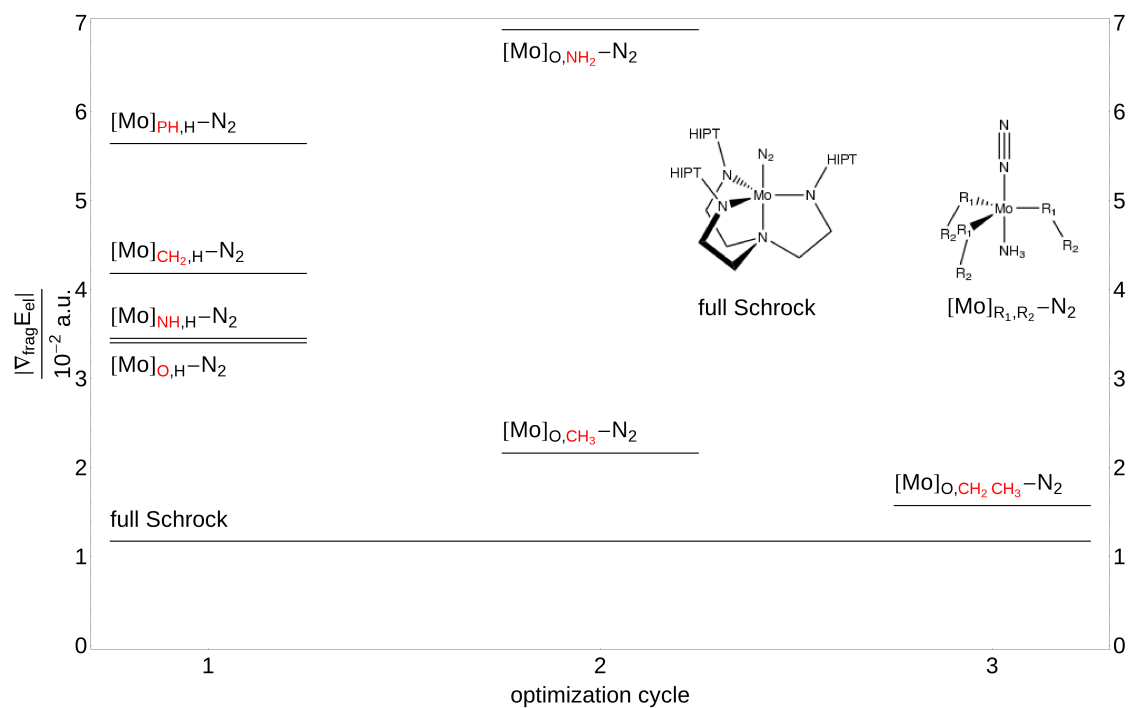


Figure 5: Stepwise construction of a nitrogen-binding transition metal complex following the concept of gradient-driven molecule construction. The data for this figure have been taken from Ref. [114] (see this reference for technical details).

design of molecules with well-defined structural features in one of its fragment structures.

## 10 Conclusion and Outlook

In this work we have aimed at providing an overview on inverse approaches in quantum chemistry for the direct rational design of molecular properties and function. The rational design of molecules and materials remains a challenging task, despite the many achievements accomplished in this field over the past decades. The most advanced contemporary methods all rely on sophisticated optimization algorithms and powerful computer hardware. Therefore, future advances in these areas will enable us to tackle increasingly complicated problems. Apart from new theoretical and algorithmic developments in inverse quantum chemistry, the possibility to carry out quantum chemical calculations (almost) instantaneously, i.e., in real time and thus interactively [117] will further enhance our capabilities (with respect to automated procedures as well as with respect to manually steered quantum-chemical design protocols [118–122]). Clearly, massive screening approaches are now feasible and help us better understand the relation between molecular structure and property. This knowledge should put us in a position to devise improved or novel inverse methods, which can be set up to serve certain well-defined design goals.

## Acknowledgments

This work has been supported by the Swiss National Science Foundation SNF (project 200020\_144458/1).

## References

- [1] *Theory and Applications of Computational Chemistry: The First Forty Years*; Dykstra, C. E.; Frenking, G.; Kim, K. S.; Scuseria, G. E., Eds.; Elsevier Science Pub-

- lishers: Amsterdam, **2005**.
- [2] Helgaker, T.; Jørgensen, P.; Olsen, J. *Molecular Electronic-Structure Theory*; John Wiley & Sons: Chichester, **2000**.
  - [3] Parr, R. G.; Yang, W. *Density-Functional Theory of Atoms and Molecules*; Oxford University Press: New York, **1989**.
  - [4] Cohen-Tannoudji, C.; Diu, B.; Laloë, F. *Quantum Mechanics*; John Wiley & Sons: New York, **1977**.
  - [5] Campman, X.; Riyanti, C. D. *Geophys. J. Int.* **2007**, *171*, 1118–1125.
  - [6] *Inverse Problems in Medical Imaging and Nondestructive Testing: Proceedings of the Conference in Oberwolfach, Federal Republic of Germany, February 4–10, 1996*; Engl, H. W.; Louis, A. K.; Rundell, W., Eds.; Springer: Wien, **1997**.
  - [7] Iyer, N.; Jayanti, S.; Lou, K.; Kalyanaraman, Y.; Ramani, K. *Computer-Aided Design* **2005**, *37*, 509–530.
  - [8] Chadán, K.; Sabatier, P. C. *Inverse Problems in Quantum Scattering Theory*, 2nd ed.; Springer: New York, **1989**.
  - [9] Karwowski, J. *Int. J. Quantum Chem.* **2009**, *109*, 2456–2463.
  - [10] Ertl, P. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 374–380.
  - [11] Bohacek, R. S.; McMartin, C.; Guida, W. C. *Med. Res. Rev.* **1996**, *16*, 3–50.
  - [12] Dobson, C. M. *Nature* **2004**, *432*, 824–828.
  - [13] Hall, G. G. *Chem. Soc. Rev.* **1973**, *2*, 21.
  - [14] Pöschel, J.; Trubowitz, E. *Inverse Spectral Theory*; Academic Press, Inc.: Orlando, **1987**.
  - [15] Note that one often encounters the English transliterations “Ambartsumian” and “Hambardzumyan” for the last name. In this paper, however, we adopt the transliteration as reported in Ambarzumian’s paper [16].

- [16] Ambarzumian, V. *Zeitschrift f. Phys.* **1929**, *53*, 690–695.
- [17] Ince, E. L. *Proc. Cam. Phil. Soc.* **1926**, *23*, 44–46.
- [18] Rayleigh, J. W. S. *The Theory of Sound*, 2nd ed.; Dover: New York, **1945**.
- [19] Borg, G. *Acta Mathematica* **1946**, *78*, 1–96.
- [20] Jansen, M.; Schön, J. C. *Nat. Mater.* **2004**, *3*, 838.
- [21] von Lilienfeld, O. A.; Lins, R. D.; Rothlisberger, U. *Phys. Rev. Lett.* **2005**, *95*, 153002.
- [22] Albritton, D. L.; Harrop, W. J.; Schmeltekopf, A. L.; Zare, R. N. *J. Mol. Spectrosc.* **1973**, *46*, 25–36.
- [23] Kirschner, S. M.; Watson, J. K. G. *J. Mol. Spectrosc.* **1974**, *51*, 321–333.
- [24] Ho, T.; Rabitz, H.; Choi, S. E.; Lester, M. I. *J. Chem. Phys.* **1995**, *102*, 2282–2285.
- [25] Zhang, D. H.; Light, J. C. *J. Chem. Phys.* **1995**, *103*, 9713–9720.
- [26] Pashov, A.; Jastrzębski, W.; Kowalczyk, P. *Comp. Phys. Comm.* **2000**, *128*, 622–634.
- [27] Kosman, W. M.; Hinze, J. *J. Mol. Spectrosc.* **1975**, *56*, 93–103.
- [28] Hamilton, I. P.; Light, J. C.; Whaley, K. B. *J. Chem. Phys.* **1986**, *85*, 5151–5157.
- [29] Wu, Q.; Zhang, J. Z. H. *Chem. Phys. Lett.* **1996**, *252*, 195–200.
- [30] Politzer, P.; Murray, J. S. *Theor. Chem. Acc.* **2002**, *108*, 134–142.
- [31] Murray, J. S.; Politzer, P. *WIREs Comput. Mol. Sci.* **2011**, *1*, 153–163.
- [32] Marder, S. R.; Beratan, D. N.; Cheng, L.-T. *Science* **1991**, *252*, 103–106.
- [33] Dalton, L. R. *J. Phys. Condens. Matter* **2003**, *15*, R897–R934.
- [34] Squitieri, E.; Paz, J. L.; Hernández, A. J. *Mol. Phys.* **2003**, *101*, 2279–2284.

- [35] Wang, T.; Moll, N.; Cho, K.; Joannopoulos, J. D. *Phys. Rev. Lett.* **1999**, *82*, 3304–3307.
- [36] Beratan, D. N. In *Materials for Nonlinear Optics: Chemical Perspectives*; Marder, S. R.; Sohn, J. E.; Stucky, G. D., Eds.; American Chemical Society: Washington, DC, **1991**, pp 89–102.
- [37] Kuhn, C.; Beratan, D. N. *J. Phys. Chem.* **1996**, *100*, 10595–10599.
- [38] Hückel, E. *Zeitschr. Elektrochem. angew. physik. Chem.* **1937**, *43*, 752–788.
- [39] Hückel, E. *Zeitschr. Elektrochem. angew. physik. Chem.* **1937**, *43*, 827–849.
- [40] Hu, X.; Beratan, D. N.; Yang, W. *Sci. China Ser. B — Chem.* **2009**, *52*, 1769–1776.
- [41] Le, T.; Epa, V. C.; Burden, F. R.; Winkler, D. A. *Chem. Rev.* **2012**, *112*, 2889–2919.
- [42] Mavromoustakos, T.; Durdagi, S.; Koukoulitsa, C.; Simcic, M.; Papadopoulos, M. G.; Hodoscek, M.; Grdadolnik, S. G. *Curr. Med. Chem.* **2011**, *18*, 2517–2530.
- [43] Karelson, M.; Lobanov, V. S. *Chem. Rev.* **1996**, *96*, 1027–1043.
- [44] Katritzky, A. R.; Lobanov, V. S.; Karelson, M. *Chem. Soc. Rev.* **1995**, *24*, 279–287.
- [45] Nantasenamat, C.; Isarankura-Na-Ayudhya, C.; Naenna, T.; Prachayasittikul, V. *EXCLI Journal* **2009**, *8*, 74–88.
- [46] Katritzky, A. R.; Kuanar, M.; Slavov, S.; Hall, C. D.; Karelson, M.; Kahn, I.; Dobchev, D. A. *Chem. Rev.* **2010**, *110*, 5714–5798.
- [47] Berhanu, W. M.; Pillai, G. G.; Oliferenko, A. A.; Katritzky, A. R. *ChemPlusChem* **2012**, *77*, 507–517.
- [48] Cros, A. F. A. Ph.D. thesis, University of Strasbourg, 1863.
- [49] Hansch, C.; Fujita, T. *J. Am. Chem. Soc.* **1964**, *86*, 1616–1626.
- [50] Hammet, L. P. *Chem. Rev.* **1935**, *17*, 125–136.

- [51] Hammett, L. P. *J. Am. Chem. Soc.* **1937**, *59*, 96–103.
- [52] These data are calculated with the program ADF, version 2010.02b [123], utilizing the BP86 exchange–correlation functional [124,125] and the TZP basis set without frozen cores [126] at all atoms. Scalar-relativistic effects were taken into account by means of the zeroth order regular approximation [127].
- [53] Schenk, S.; Reiher, M. *Inorg. Chem.* **2009**, *48*, 1638–1648.
- [54] Zheng, W.; Cho, S. J.; Tropsha, A. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 251–258.
- [55] Cho, S. J.; Zheng, W.; Tropsha, A. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 259–268.
- [56] Kirkpatrick, S.; Gelatt, Jr., C. D.; Vecchi, M. P. *Science* **1983**, *220*, 671–680.
- [57] Mühlenbein, H.; Gorges-Schleuter, M.; Krämer, O. *Parallel Comput.* **1988**, *7*, 65–85.
- [58] Metropolis, N.; Ulam, S. *J. Am. Stat. Assoc.* **1949**, *44*, 335–341.
- [59] Schön, J. C.; Jansen, M. *Angew. Chem. Int. Ed. Engl.* **1996**, *35*, 1286–1304.
- [60] Jansen, M. *Angew. Chem. Int. Ed.* **2002**, *41*, 3746–3766.
- [61] Putz, H.; Schön, J. C.; Jansen, M. *Ber. Bunsenges. Phys. Chem.* **1995**, *99*, 1148–1153.
- [62] Schön, J. C.; Jansen, M. *Comput. Mater. Sci.* **1995**, *4*, 43–58.
- [63] Werner, J. H.; Kolodinski, S.; Queisser, H. J. *Phys. Rev. Lett.* **1994**, *72*, 3851–3854.
- [64] Franceschetti, A.; Zunger, A. *Nature* **1999**, *402*, 60–63.
- [65] Silverman, A.; Zunger, A.; Kalish, R.; Adler, J. *Phys. Rev. B* **1995**, *51*, 10795–10816.
- [66] Wang, L.-W.; Zunger, A. *Phys. Rev. B* **1995**, *51*, 17398–17416.
- [67] Wang, L.-W.; Zunger, A. *J. Chem. Phys.* **1994**, *100*, 2394–2397.
- [68] Piquini, P.; Graf, P. A.; Zunger, A. *Phys. Rev. Lett.* **2008**, *100*, 186403.

- [69] Kim, K.; Graf, P. A.; Jones, W. B. *J. Comput. Phys.* **2005**, *208*, 735–760.
- [70] Dudiy, S. V.; Zunger, A. *Phys. Rev. Lett.* **2006**, *97*, 046401–1.
- [71] Jóhannesson, G. H.; Bligaard, T.; Ruban, A. V.; Skriver, H. L.; Jacobsen, K. W.; Nørskov, J. K. *Phys. Rev. Lett.* **2002**, *88*, 255506–1.
- [72] Nørskov, J. K.; Bligaard, T.; Rossmeisl, J.; Christensen, C. H. *Nature Chem.* **2009**, *1*, 37–46.
- [73] Hautier, G.; Jain, A.; Ong, S. P. *J. Mater. Sci.* **2012**, *47*, 7317–7340.
- [74] Marzari, N. *MRS Bulletin* **2006**, *31*, 681–687.
- [75] Ceder, G. *MRS Bulletin* **2010**, *35*, 693–701.
- [76] Chan, M. K. Y.; Ceder, G. *Phys. Rev. Lett.* **2010**, *105*, 196403.
- [77] Jain, A.; Ong, S. P.; Hautier, G.; Chen, W.; Richards, W. D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G.; Persson, K. A. *Appl. Phys. Lett. Mater.* **2013**, *1*, 011002.
- [78] Hautier, G.; Fischer, C.; Ehrlacher, V.; Jain, A.; Ceder, G. *Inorg. Chem.* **2011**, *50*, 656–663.
- [79] Struebing, H.; Ganase, Z.; Karamertzanis, P. G.; Siougkrou1, E.; Haycock, P.; Piccione, P. M.; Armstrong, A.; Galindo1, A.; Adjiman, C. S. *Nature Chem.* **2013**, *5*, 952–957.
- [80] Truhlar, D. G. *Nature Chem.* **2013**, *5*, 902–903.
- [81] Wang, M.; Hu, X.; Beratan, D. N.; Yang, W. *J. Am. Chem. Soc.* **2006**, *128*, 3228–3232.
- [82] Hohenberg, P.; Kohn, W. *Phys. Rev.* **1964**, *136*, 864–871.
- [83] Yang, W.; Ayers, P. W.; Wu, Q. *Phys. Rev. Lett.* **2004**, *92*, 146404.
- [84] Bendsøe, M. P.; Kikuchi, N. *Comp. Methods Appl. Mech. Eng.* **1988**, *71*, 197–224.



- [85] Keinan, S.; Therien, M. J.; Beratan, D. N.; Yang, W. *J. Phys. Chem. A* **2008**, *112*, 12203–12207.
- [86] d’Avezac, M.; Zunger, A. *J. Phys. Condens. Matter*. **2007**, *19*, 402201.
- [87] Keinan, S.; Hu, X.; Beratan, D. N.; Yang, W. *J. Phys. Chem. A* **2007**, *111*, 176–181.
- [88] Hu, X.; Beratan, D. N.; Yang, W. *J. Chem. Phys.* **2008**, *129*, 064102.
- [89] Xiao, D.; Yang, W.; Beratan, D. N. *J. Chem. Phys.* **2008**, *129*, 044106.
- [90] Pearl, J. *Heuristics: Intelligent Search Strategies for Computer Problem Solving*; Addison-Wesley: Reading, Massachusetts, **1984**.
- [91] Balamurugan, D.; Yang, W.; Beratan, D. N. *J. Chem. Phys.* **2008**, *129*, 174105.
- [92] Cohen, A. J.; Mori-Sánchez, P.; Yang, W. *Science* **2008**, *321*, 792–794.
- [93] von Lilienfeld, O. A. *Int. J. Quantum Chem.* **2013**, *113*, 1676–1689.
- [94] von Lilienfeld, O. A.; Tuckerman, M. E. *J. Chem. Phys.* **2006**, *125*, 154104.
- [95] von Lilienfeld, O. A.; Tuckerman, M. E. *J. Chem. Theory Comput.* **2007**, *3*, 1083–1090.
- [96] von Lilienfeld, O. A. *J. Chem. Phys.* **2009**, *131*, 164102.
- [97] Sheppard, D.; Henkelman, G.; von Lilienfeld, O. A. *J. Chem. Phys.* **2010**, *133*, 084104.
- [98] Marcon, V.; von Lilienfeld, O. A.; Andrienko, D. *J. Chem. Phys.* **2007**, *127*, 064305.
- [99] Herrmann, C.; Neugebauer, J.; Reiher, M. *New J. Chem.* **2007**, *31*, 818–831.
- [100] Reiher, M.; Neugebauer, J. *J. Chem. Phys.* **2003**, *118*, 1634–1641.
- [101] Kiewisch, K.; Lubner, S.; Neugebauer, J.; Reiher, M. *Chimia* **2009**, *63*, 270–274.
- [102] Kiewisch, K.; Neugebauer, J.; Reiher, M. *J. Chem. Phys.* **2008**, *129*, 204103.

- [103] Lubner, S.; Neugebauer, J.; Reiher, M. *J. Chem. Phys.* **2009**, *130*, 064105.
- [104] Lubner, S.; Reiher, M. *ChemPhysChem* **2009**, *10*, 2049–2057.
- [105] Neugebauer, J.; Reiher, M.; Kind, C.; Hess, B. A. *J. Comput. Chem.* **2002**, *23*, 895.
- [106] Reiher, M.; Neugebauer, J. *Phys. Chem. Chem. Phys.* **2004**, *6*, 4621–4629.
- [107] Neugebauer, J.; Reiher, M. *J. Comput. Chem.* **2004**, *25*, 587–597.
- [108] Neugebauer, J.; Reiher, M. *J. Phys. Chem. A* **2004**, *108*, 2053–2061.
- [109] Herrmann, C.; Reiher, M. *Surf. Science* **2006**, *600*, 1891–1900.
- [110] Adler, T. B.; Borho, N.; Reiher, M.; Suhm, M. A. *Angew. Chem.* **2006**, *118*, 3518–3523.
- [111] Herrmann, C.; Neugebauer, J.; Reiher, M. *J. Comput. Chem* **2008**, *29*, 2460–2470.
- [112] Kovyrshin, A.; Neugebauer, J. *J. Chem. Phys.* **2010**, *133*, 174114.
- [113] Kovyrshin, A.; Neugebauer, J. *Chem. Phys.* **2011**, *391*, 147–156.
- [114] Weymuth, T.; Reiher, M. *Int. J. Quantum Chem.* **2014**, *114*, 838–850.
- [115] Weymuth, T.; Reiher, M. *MRS Proceedings* **2013**, *1524*, DOI: 10.1557/opl.2012.1764.
- [116] Reiher, M.; Kirchner, B.; Hutter, J.; Sellmann, D.; Hess, B. A. *Chem. Eur. J.* **2004**, *10*, 4443–4453.
- [117] Haag, M. P.; Reiher, M. *Int. J. Quantum Chem.* **2013**, *113*, 8–20.
- [118] Marti, K. H.; Reiher, M. *J. Comput. Chem.* **2009**, *30*, 2010–2020.
- [119] Haag, M. P.; Marti, K. H.; Reiher, M. *ChemPhysChem* **2011**, *12*, 3204–3213.
- [120] Bosson, M.; Grudinin, S.; Bouju, X.; Redon, S. *J. Comput. Phys.* **2011**, *231*, 2581–2598.

- [121] Bosson, M.; Richard, C.; Plet, A.; Grudinin, S.; Redon, S. *J. Comput. Chem.* **2012**, *33*, 779–790.
- [122] Haag, M. P.; Reiher, M. *Faraday Discuss.* **2014**, *169*, DOI: 10.1039/C4FD00021H
- [123] te Velde, G.; Bickelhaupt, F. M.; van Gisbergen, S. J. A.; Guerra, C. F.; Baerends, E. J.; Snijders, J. G.; Ziegler, T. *J. Comput. Chem.* **2001**, *22*, 931.
- [124] Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098.
- [125] Perdew, J. P. *Phys. Rev. B* **1986**, *33*, 8822.
- [126] van Lenthe, E.; Baerends, E. J. *J. Comput. Chem.* **2003**, *24*, 1142.
- [127] van Lenthe, E.; Ehlers, A. E.; Baerends, E. J. *J. Chem. Phys.* **1999**, *110*, 8943.